SONDAGE DANS UNE POPULATION ANIMALE : ESTIMATION DU TAUX D'INFECTION DES CHEPTELS

M. ELOIT et M.A. KOUTCHOUKALI**

==========

RESUME : Les auteurs présentent les principes permettant d'établir un protocole de sondage visant à estimer le taux d'infection des cheptels vis-à-vis d'une maladie infectieuse au sein d'une population animale. Le nombre de cheptels à étudier ainsi que le nombre d'animaux à tester par cheptel sont évoqués en fonction de divers paramètres.

SUMMARY : The authors present the principles leading to the sampling rules to be used when looking for the infection rate of herds in case of an infectious disease within an animal population. The number of herds to study, as well as the number of animals to test in each herd are presented, following different situations.

***

La décision de mise en oeuvre d'une prophylaxie collective d'une maladie infectieuse, et la définition de ses modalités demandent d'apprécier au préalable le taux d'infection des cheptels au sein d'une population animale.

Ce document a pour but de mettre à la disposition des utilisateurs les éléments permettant de bâtir un protocole de sondage destiné à estimer le taux d'infection des cheptels.

L'élaboration de ce protocole conduit à répondre à 3 questions :

. Comment choisir les cheptels à contrôler ?
. Quel nombre de cheptels contrôler ?
. Quelles sont les modalités d'appréciation de l'état d'infection d'un cheptel ?

Nous décrirons ce qui a trait au sondage aléatoire simple, en laissant de côté les techniques de stratification et de quotas.

---

* *Ecole nationale vétérinaire d'Alfort - Chaire de maladies contagieuses - 94704 Maisons-Alfort cedex.*

I. CHOIX DES CHEPTELS A CONTROLER

Cet aspect pose le problème des modalités du tirage au sort qui doit se rapprocher le plus possible d'un modèle complètement aléatoire. Le postulat est qu'un tirage au hasard prend naturellement en compte toutes les variables qui peuvent modifier le taux d'infection à estimer, alors qu'un choix "raisonné" des exploitations conduit inévitablement à en privilègier certaines a priori.

Lorsque l'on possède un fichier exhaustif des exploitations de la population à sonder, il est préférable de recourir à une table de nombres au hasard (tableau I). D'autre part, certaines calculatrices, ainsi que les micro-ordinateurs, possèdent une fonction permettant de générer des nombres au hasard.

L'absence de fichier exhaustif des exploitations conduit à imaginer dans chaque cas une modalité de tirage au sort qui respecte au mieux le caractère aléatoire du choix. A titre d'exemple, signalons quelques biais possibles :

- exploitations au contrôle laitier (non représentatif du niveau zootechnique et sanitaire moyen) ;
- adhérents d'un même groupement de producteurs (contacts entre ces exploitations sans doute plus fréquents qu'avec les exploitations appartenant à un autre groupement) ;
- clientèle d'un vétérinaire (localisation géographique).

En conclusion de ce bref chapitre, il est nécessaire de souligner que le soin et les efforts correspondant au choix des exploitations conditionnent très largement la fiabilité du sondage. En particulier, l'estimation du nombre de cheptels à contrôler (chapitre suivant) est basée sur un sondage de caractère aléatoire. Enfin, un sondage établi sur la base d'un choix d'exploitations que l'on sait biaisé est potentiellement facteur d'erreurs car les chiffres obtenus, après diffusion, prennent rapidement une valeur absolue, alors que les réserves qui les accompagnent sombrent généralement dans l'oubli.

II. NOMBRE DE CHEPTELS A CONTROLER

Le problème à résoudre est de connaître le nombre de cheptels à contrôler pour estimer le taux d'infection des cheptels d'une population, avec une précision préalablement définie.

1. Bases théoriques

La variance d'une proportion estimée (p) sur un échantillon de taille n issu d'une population de taille infinie est :

$$\text{variance } (p) = \frac{p\,(1 - p)}{n} \qquad (1)$$

Quand la taille de l'échantillon ne peut être considérée comme négligeable vis-à-vis de celle de la population, cette variance doit être corrigée par un facteur dit "facteur d'exhaustivité" :

$$\text{variance } (p) = \frac{N - n}{N} \; \frac{p\,(1 - p)}{n} \qquad (2)$$

avec N = taille de la population.

Tableau I : Table de nombres au hasard. Pour obtenir une suite de nombres aléatoires, il suffit de suivre la table selon un ordre choisi par l'utilisateur.

Exemple : Nombre à 3 chiffres

- On peut choisir de prendre les 3 derniers chiffres de chaque nombre en suivant la table colonne par colonne ; on obtient la série : 881, 912, 952, 937, 718...
- On peut également choisir de prendre les 3 chiffres du milieu de chaque nombre en suivant la table ligne par ligne ; on obtient la série : 688, 658, 896, 607, 918...

Table de nombres au hasard

06881	56588	88965	46072	59182	56878	73801	34975	09952
36912	00172	65326	30491	29431	45754	38950	82193	08100
34952	07966	28047	29892	92667	67055	62348	29600	20655
82937	14987	64121	75335	70393	71237	68907	34100	31364
66718	57759	11646	28582	17746	21163	20693	97161	60862
78931	78939	98988	48471	76708	90936	91995	02728	81340
55704	05132	64978	78997	01260	06295	69334	73955	42208
86899	92257	67149	06277	46811	71722	18330	99968	70211
43258	49277	36261	37037	48348	28096	52143	79139	53799
79779	27597	02987	08812	35097	24990	75781	38821	70858
34475	24722	62266	32415	95047	50331	27223	79523	85599
87855	81968	31132	44535	19247	27304	36732	84287	98688
60188	58144	80971	45390	13431	40061	43008	73826	36744
57475	92686	98674	71524	75554	19210	96969	75078	15972
47413	22826	87099	51745	98290	53876	75809	22095	62118
13256	53481	06541	75270	69146	17497	07218	94613	13722
96692	85279	71033	70038	91547	87876	95221	20359	74041
12435	43207	36897	97584	97223	24541	65438	08015	23903
31948	83681	04895	83471	44957	80229	97907	16651	37744
03857	87403	14987	73345	08864	02924	16504	29003	12287
23963	35532	18924	16152	56268	54025	43179	01528	18505
93058	22447	70290	47410	26659	89476	78079	27303	66516
39864	28728	89225	51604	91171	91589	59101	59284	65756
08210	05059	66007	38287	98560	03075	90849	71739	32817
27271	60035	81718	80042	42121	72344	93751	91178	86321
49714	36622	29297	19633	69616	01776	93564	55168	91618
58674	64079	84747	35031	24636	56769	35847	32716	19039
26579	24847	79085	85220	79467	93508	21885	28899	92382
18336	84469	98574	70016	36441	48433	41671	10895	96530
44457	94464	72971	77096	47440	73187	44242	08446	88814
10369	52908	77153	95194	82923	75649	09738	67852	64310
33470	44759	86877	27021	61875	17695	76324	91034	51960
62164	57763	68794	85145	76870	47148	90046	23111	46909
23801	80067	91140	88369	39816	35799	91102	65191	25755
19170	64616	50987	53386	87818	45050	91490	60354	24362

Quand $n \ll N$, les formules (1) et (2) apparaissent bien comme équivalentes. L'intervalle de confiance de la proportion estimée est alors de :

$$p \pm \varepsilon \sqrt{\frac{N - n}{N} \quad \frac{p\,(1 - p)}{n}} \qquad (3)$$

avec ε = variable normale réduite au risque choisi.

La précision absolue sur la proportion estimée est donc :

$$PA = \varepsilon \sqrt{\frac{N - n}{N} \quad \frac{p\,(1 - p)}{n}} \qquad (4)$$

et la précision relative est :

$$PR = \frac{PA}{p} = \varepsilon \cdot \left(\sqrt{\frac{N - n}{N} \quad \frac{p\,(1 - p)}{n}}\right) \cdot \left(\frac{1}{p}\right)$$

$$PR = \varepsilon \sqrt{\frac{N - n}{N} \quad \frac{(1 - p)}{np}} \qquad (5)$$

On peut donc estimer n à partir de N, de p et de PA en utilisant la formule (4).

$$(PA)^2 = \varepsilon^2 \left(1 - \frac{n}{N}\right) \; p \; \frac{(1 - p)}{n}$$

$$(PA)^2 = \varepsilon^2 \left(\frac{1}{n} - \frac{1}{N}\right) p\,(1 - p)$$

$$\frac{(PA)^2}{\varepsilon^2\, p\,(1 - p)} + \frac{1}{N} = \frac{1}{n}$$

$$\text{d'où} \qquad n = \frac{\varepsilon^2 N \; p\,(1 - p)}{N\,(PA)^2 + \varepsilon^2\, p\,(1 - p)} \qquad (6)$$

De la même façon, on peut estimer n à partir de N, de p et de PR en utilisant la formule (5) :

$$(PR)^2 = \varepsilon^2 \left(1 - \frac{n}{N}\right) \left(\frac{1 - p}{np}\right)$$

$$(PR)^2 = \varepsilon^2 \left(\frac{1}{n} - \frac{1}{N}\right) \left(\frac{1 - p}{p}\right)$$

$$\frac{(PR)^2\, p}{\varepsilon^2\,(1 - p)} + \frac{1}{N} = \frac{1}{n}$$

$$\text{d'où} \qquad n = \frac{\varepsilon^2 N\,(1 - p)}{N\,(PR)^2\, p + \varepsilon^2\,(1 - p)} \qquad (7)$$

On passe également de (6) à (7) en posant dans (6):PA = PR x p.

N.B. : Ces formules sont valides si np et n(1 - p) sont supérieurs à la valeur 5.

En résumé, on peut donc estimer n (nombre de cheptels à contrôler) en fixant au préalable 3 paramètres :

. la taille de la population (N)
. le taux d'infection approximatif que l'on cherche à mettre en évidence,
. la précision, relative ou absolue, que l'on attend du sondage.

Il faut noter que, pour une précision absolue donnée, l'effectif à contrôler est d'autant plus important que le taux d'infection à estimer est proche de 50 %. D'autre part, pour une précision relative donnée, l'effectif à contrôler est d'autant plus important que le taux d'infection à estimer est faible. En corollaire, en fonction de l'expression de la précision choisie, il faudra fournir comme paramètre pour le taux d'infection à estimer une valeur minimale (précision relative) ou maximale (précision absolue).

Il parait néanmoins préférable d'exprimer la précision en valeur relative : en effet, une précision absolue de 5 % n'a pas la même signification si elle correspond à un taux d'infection de 50 % ou de 5 %, alors qu'une précision relative de 20 % définit une précision absolue de 10 % pour un taux d'infection de 50 % et de 1 % pour un taux d'infection de 5 % (respectivement 50 % x 0,2 et 5 % X 0,2). C'est pourquoi dans la suite de l'exposé, nous présenterons des tableaux et schémas en fonction de valeurs de précision relative.

Notons enfin que la formule (4) permet de calculer, une fois les résultats du sondage connus, la précision absolue sur le taux d'infection estimé. Cette précision calculée est la meilleure que l'on puisse espérer, ne serait-ce qu'en raison du caractère fortement hétérogène de la population sondée.

2. Valeurs calculées

. Le tableau II et la figure 1 présentent le nombre de cheptels à contrôler en fonction :

- du nombre de cheptels de la population,
- du taux d'infection minimal que l'on cherche à estimer,
- d'une précision relative souhaitée de 20 %.

. Le tableau III et la figure 2 présentent les mêmes données pour une précision relative de 33 %.

. Le tableau IV et la figure 3 sont équivalents aux précédents pour une précision relative de 50 %.

Dans tous les cas le seuil de risque choisi a été de 5 %.

III. MODALITES D'APPRECIATION DE L'ETAT D'INFECTION D'UN CHEPTEL

Pour des raisons de coût, il est souvent impossible de tester tous les animaux d'un cheptel. Si l'on écarte l'alternative d'utiliser des prélèvements de grand mélange (diagnostic sérologique), il est possible de ne prélever qu'un certain nombre d'animaux dans l'exploitation en fonction de critères exposés dans les lignes suivantes.

Tableau II : Nombre de cheptels à contrôler en fonction du taux d'infection minimal à estimer et du nombre de cheptels total (précision relative 20 %).

Nombre de Cheptels	Taux d'Infection (Précision relative:20%)							
	2%	5%	10%	15%	20%	30%	40%	50%
100	98	95	90	85	80	70	60	50
250	238	221	196	174	154	121	94	72
500	454	396	322	266	223	160	116	84
1000	831	656	474	362	286	190	131	91
2500	1656	1080	662	462	345	214	142	97
5000	2475	1377	763	509	371	223	146	99
10000	3289	1597	826	537	385	229	148	100
15000	3694	1687	850	547	390	230	149	100
20000	3936	1736	862	552	393	231	149	100
30000	4213	1787	874	557	395	232	150	100

Figure 1 : Nombre de cheptels à contrôler en fonction du taux d'infection minimal à estimer et du nombre de cheptels total (précision relative 20 %). Les échelles sont logarithmiques.

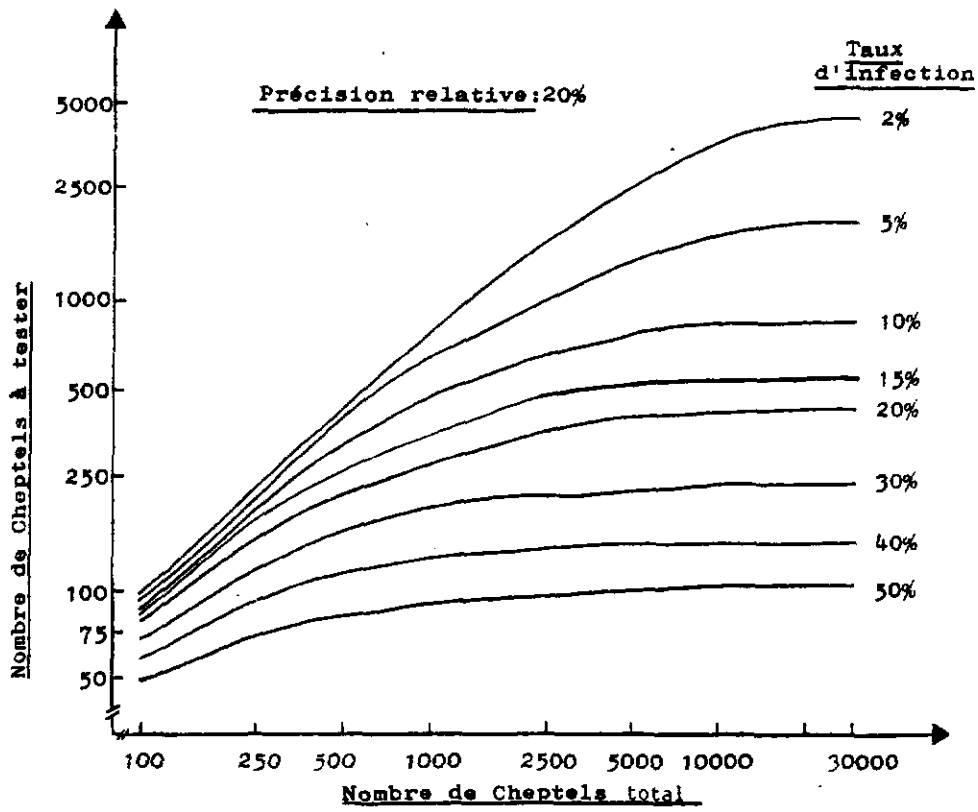


Tableau III : Nombre de cheptels à contrôler en fonction du taux d'infection minimal à estimer et du nombre de cheptels total (précision relative 33 %).

Nombre de Cheptels	Taux d'Infection (Précision relative:33%)							
	2%	5%	10%	15%	20%	30%	40%	50%
100	95	88	77	68	60	47	36	27
250	220	185	143	114	93	64	46	33
500	392	292	200	147	114	74	50	35
1000	643	412	249	173	129	79	53	36
2500	1047	546	292	193	139	83	54	37
5000	1324	613	311	200	143	85	55	37
10000	1526	653	320	204	145	85	55	37
15000	1607	667	324	206	146	86	55	37
20000	1652	675	326	206	146	86	55	37
30000	1698	683	327	207	147	86	55	37

Figure 2 : Nombre de cheptels à contrôler en fonction du taux d'infection minimal à estimer et du nombre de cheptels total (précision relative 33 %). Les échelles sont logarithmiques.

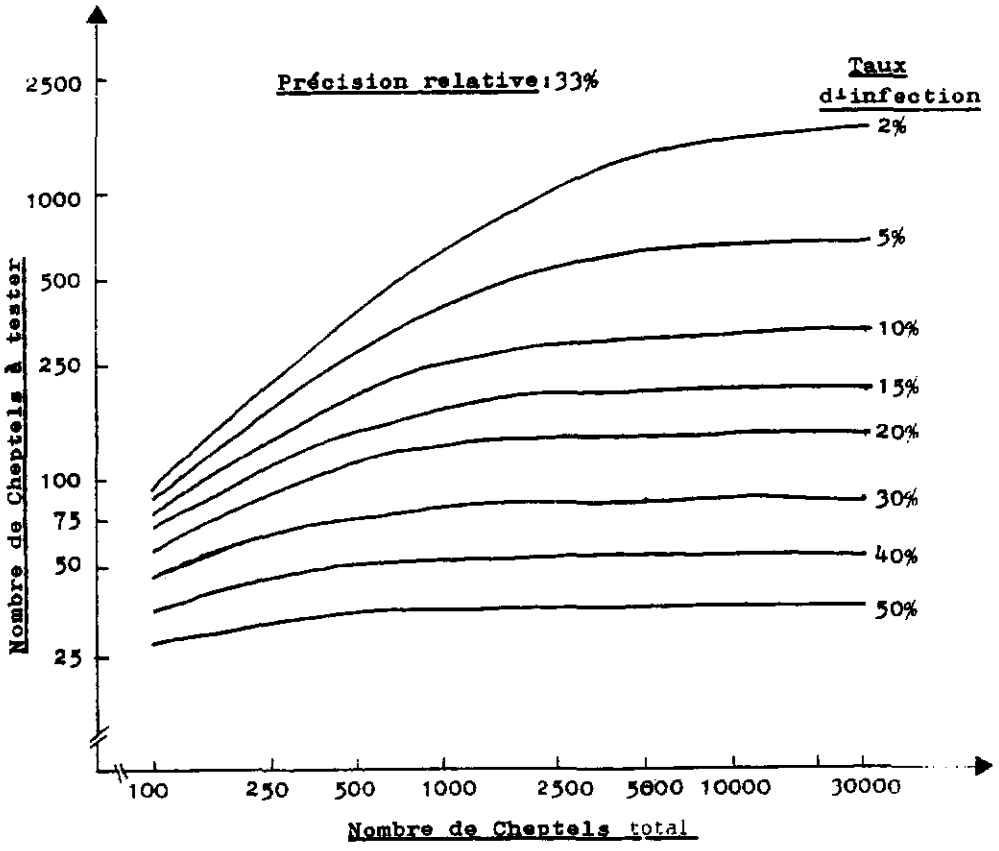
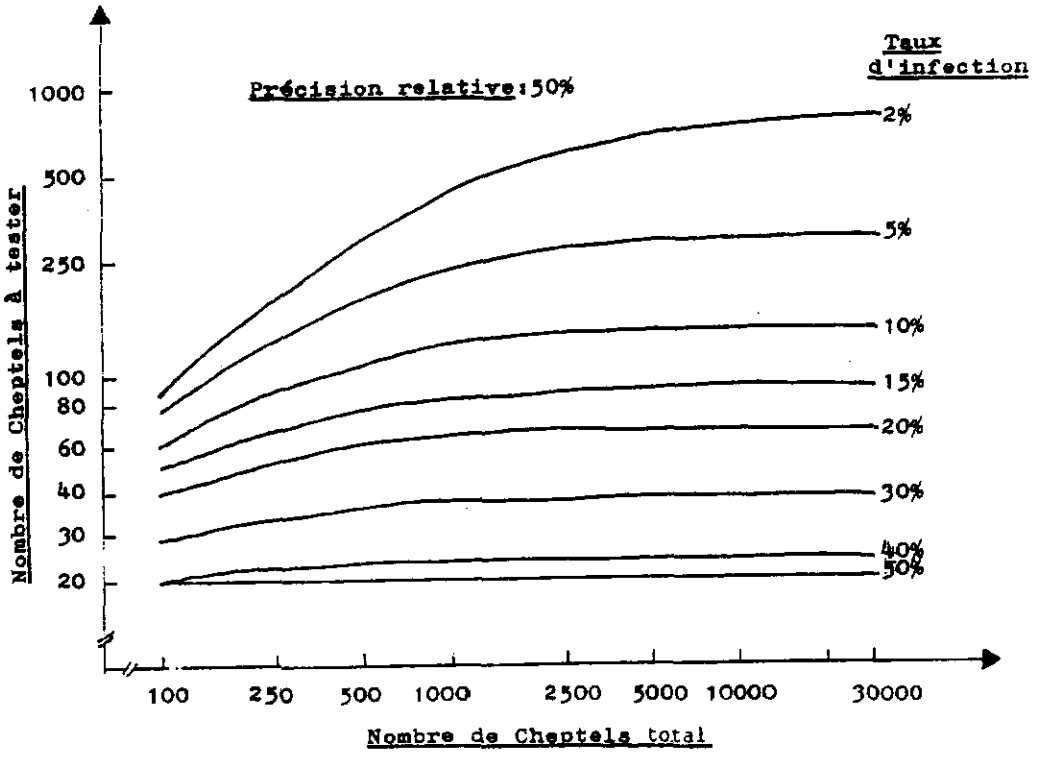


Tableau IV : Nombre de cheptels à contrôler en fonction du taux d'infection minimal à estimer et du nombre de cheptels total (précision relative 50 %).

Nombre de Cheptels	Taux d'Infection (Précison relative:50%)							
	2%	5%	10%	15%	20%	30%	40%	50%
100	89	76	60	48	40	28	20	20
250	190	138	92	67	51	33	22	20
500	306	190	112	77	57	35	23	20
1000	440	234	126	84	61	36	24	20
2500	597	272	137	88	63	37	24	20
5000	678	287	140	90	64	38	24	20
10000	728	296	142	90	64	38	24	20
15000	746	298	143	91	64	38	24	20
20000	755	300	143	91	64	38	24	20
30000	765	301	144	91	64	38	24	20

Figure 3 : Nombre de cheptels à contrôler en fonction du taux d'infection minimal à estimer et du nombre de cheptels total (précision relative 50 %). Les échelles sont logarithmiques.



1. Bases théoriques

Le problème posé est le suivant : dans un cheptel de N animaux dont une proportion p est infectée, combien faut-il contrôler d'animaux pour trouver au moins un positif (et ce, pour une probabilité donnée). La mise en évidence d'un animal infecté suffit pour considérer le cheptel comme infecté.

Sans démonstration, la formule approchée est la suivante :

$$n = (N - \frac{pN}{2} + \frac{1}{2})\ (1 - \alpha^{1/pN}) \qquad (8)$$

avec n = nombre d'animaux à contrôler
p = proportion d'animaux infectés
N = nombre d'animaux du cheptel
α = seuil de risque choisi.

Il est intuitivement évident et vérifiable par le calcul que le nombre d'animaux à tester est d'autant plus grand que le taux d'infection des animaux du cheptel est faible. Autrement dit, dans le cadre d'une enquête, il est possible de ne prélever qu'une fraction de chaque cheptel, à condition de définir au préalable le taux d'infection minimal des animaux au sein d'un cheptel au dessus duquel on refuse de ne pas détecter l'infection de l'effectif (et ceci pour un risque donné). Il reste néanmoins que des cheptels plus faiblement infectés seront repérés mais avec une probabilité plus faible.

A partir de la formule (8) on peut tirer α qui est le risque de ne pas détecter un cheptel infecté, connaissant le nombre d'animaux du cheptel, le taux d'infection des animaux, et le nombre d'animaux contrôlés.

$$n = (N - \frac{pN}{2} + \frac{1}{2})\ (1 - \alpha^{1/pN}) \qquad (8)$$

d'où

$$1 - \frac{n}{N - \frac{pN}{2} + \frac{1}{2}} = \alpha^{1/pN}$$

d'où

$$\log\ (1 - \frac{n}{N - \frac{pN}{2} + \frac{1}{2}}) = \frac{1}{pN} \log\ (\alpha\ )$$

d'où

$$\log\ (1 - \frac{n}{N - \frac{pN}{2} + \frac{1}{2}})\ pN = \log\ (\alpha\ )$$

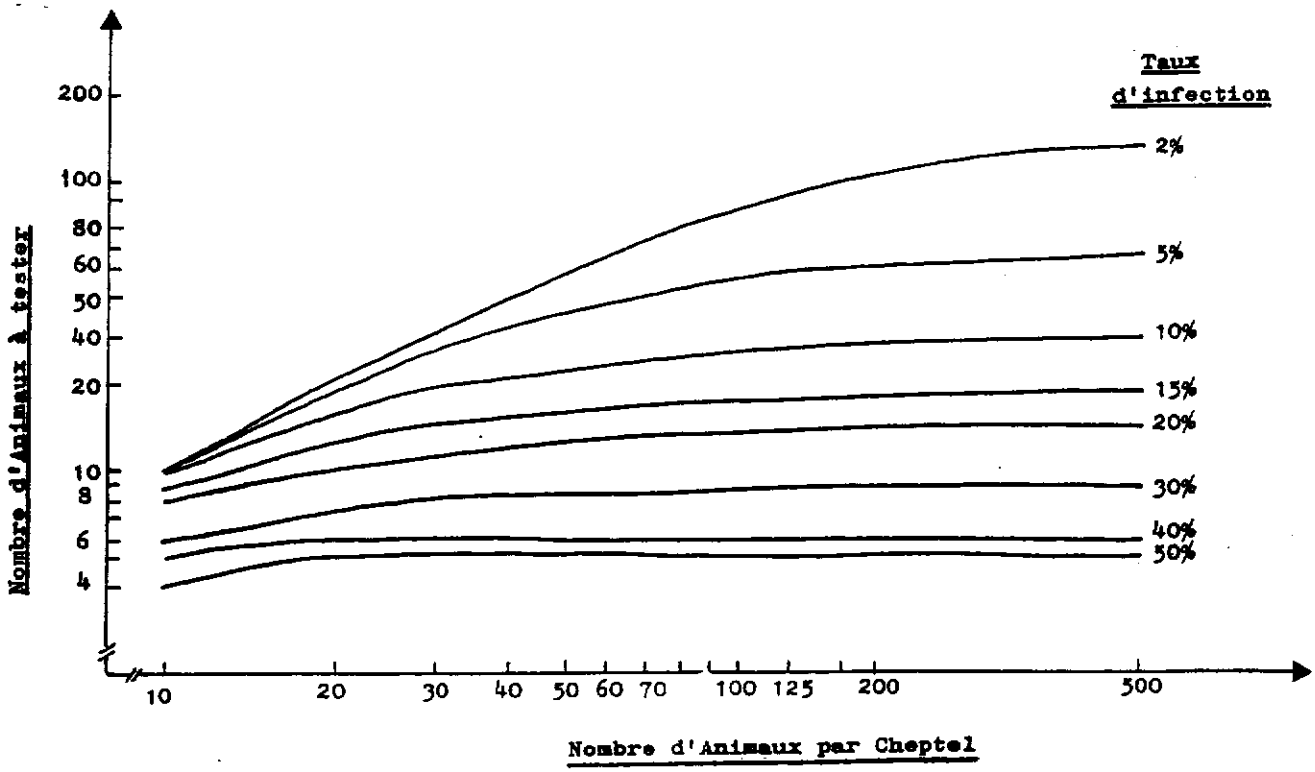
Donc

$$\boxed{\alpha = 10^{\log\ (1 - \frac{n}{N - \frac{pN}{2} + \frac{1}{2}})\ pN}} \qquad (9)$$

Tableau V : Nombre d'animaux à contrôler dans un cheptel en fonction du nombre d'animaux du cheptel et du taux minimal d'infection des animaux pour lequel on souhaite repérer l'état d'infection du cheptel (probabilité : 95 %).

Nombre d'Animaux par Cheptel	Taux d'Infection 2%	5%	10%	15%	20%	30%	40%	50%
10	10	10	10	9	8	6	5	4
20	20	19	16	13	10	7	6	5
30	30	26	19	14	11	8	6	5
40	40	31	21	15	12	8	6	5
50	48	35	22	16	12	8	6	5
60	55	38	23	16	13	8	6	5
70	62	40	24	17	13	8	6	5
80	68	42	24	17	13	9	6	5
90	73	43	25	17	13	9	6	5
100	78	45	25	17	13	9	6	5
125	87	47	26	18	13	9	6	5
150	95	49	26	18	13	9	6	5
175	100	50	27	18	13	9	6	5
200	105	51	27	18	14	9	6	5
500	129	56	28	19	14	9	6	5

Figure 4 : Nombre d'animaux à contrôler dans un cheptel en fonction du nombre d'animaux du cheptel et du taux minimal d'infection des animaux pour lequel on souhaite repérer l'état d'infection du cheptel (probabilité : 95 %). Les échelles sont logarithmiques.



2. Valeurs calculées

Le tableau V et la figure 4 présentent le nombre d'animaux à tester dans chaque cheptel en fonction :

- du nombre d'animaux du cheptel,
- du taux minimal d'infection des animaux au dessus duquel on refuse que le risque de ne pas détecter l'infection de l'effectif soit supérieur à 5 %.

Dans une enquête, ce taux doit être choisi une fois pour toutes.

Le tableau VI et la figure 5 montrent l'évolution du risque de ne pas détecter un cheptel infecté lorsque le taux d'infection du cheptel est inférieur à celui retenu comme seuil dans les lignes précédentes.

Dans le cadre d'examens sérologiques, il est possible, pour diminuer le coût total, d'effectuer les réactions au laboratoire sur des petits mélanges de sérums : le nombre de sérums mélangés doit tenir compte de la sensibilité de la technique confrontée au fait que le plus souvent, dans les exploitations dont le taux d'infection est proche de celui défini dans les lignes précédentes, un seul sérum se révèlera positif.

Tableau VI : Probabilité (en %) de détection de l'infection d'un cheptel en fonction :

- de son taux d'infection
- d'un nombre d'animaux prélevés permettant de repérer dans 95 % des cas l'état d'infection d'un cheptel pour les taux suivants : 2 %, 5 %, 10 %, 15 %, 20 %, 30 %, 40 %, 50 %.

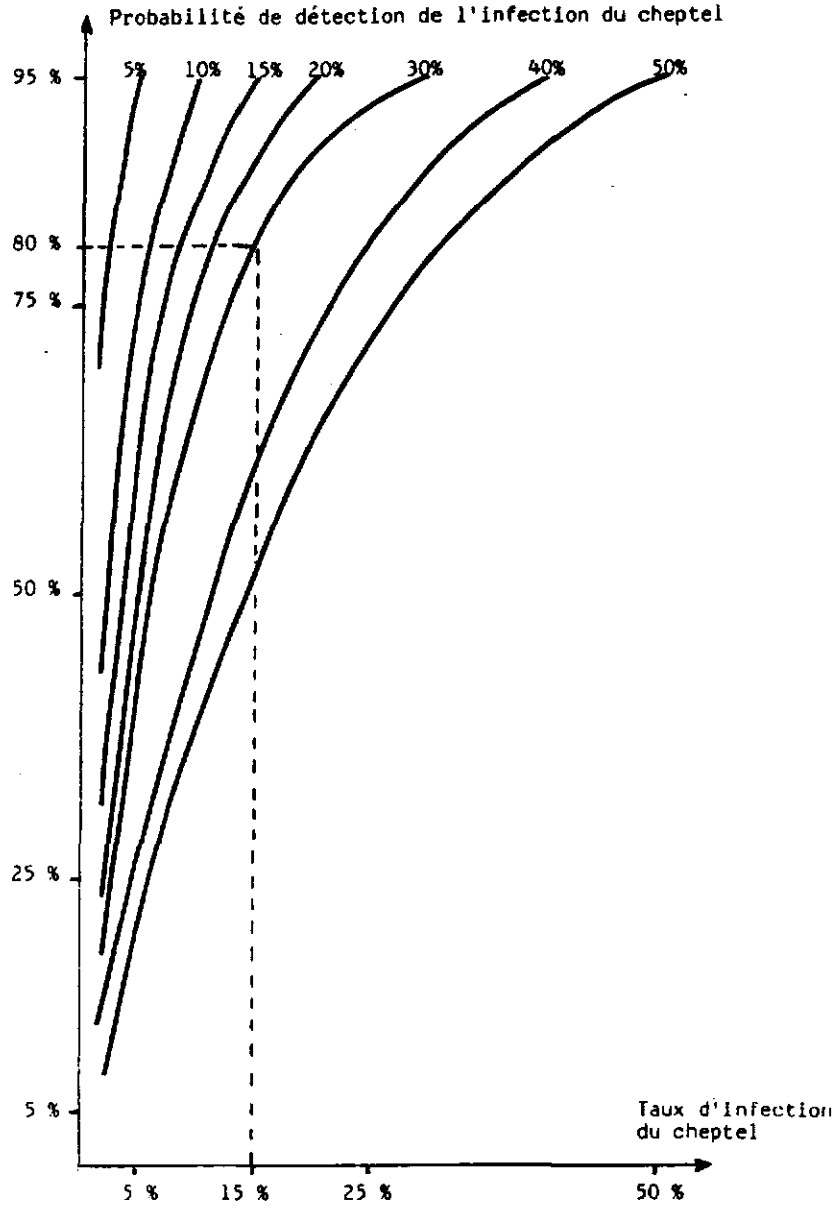
Ainsi un choix du nombre d'animaux prélevés permettant de repérer dans 95 % des cas un cheptel infecté à 30 % (colonne 30 %), sera suffisant pour mettre en évidence dans 80 % des cas les cheptels infectés à 15 % (ligne 15 %).

Taux d'infection choisi comme critère de choix du nombre d'animaux à prélever		2 %	5 %	10 %	15 %	20 %	30 %	40 %	50 %
Taux réel d'infection du cheptel	2 %	95	69	43	31	23	18	11	8
	5 %		95	76	61	49	40	26	20
	10 %			95	85	75	65	47	37
	15 %				95	88	80	62	51
	20 %					95	89	73	63
	30 %						95	88	79
	40 %							95	89
	50 %								95

Figure 5 : Probabilité (en %) de détection de l'infection d'un cheptel en fonction :

- de son taux d'infection
- d'un nombre d'animaux prélevés permettant de repérer dans 95 % des cas l'état d'infection d'un cheptel pour les taux suivants : 5 %, 10 %, 15 %, 20 %, 30 %, 40 %, 50 %.

En reprenant l'exemple du tableau VI, on peut visualier que le choix d'un nombre d'animaux prélevés permettant de repérer dans 95 % un cheptel infecté à 30 % (courbe 30 %) sera suffisant pour mettre en évidence dans 80 % des cas les cheptels infectés à 15 % (projection sur l'axe des ordonnés de l'intersection de la valeur 15 % en abscisse avec la courbe 30 %.



IV. EXEMPLE

L'exemple suivant permet de faire la synthèse des informations précédentes.

Dans un département, on souhaite faire un sondage sérologique concernant une maladie infectieuse donnée. Il existe 10.000 cheptels dans le département avec un fichier géré sur ordinateur.

Des renseignements antérieurs indiquent que le taux d'infection des cheptels est sans doute supérieur à 10 %. Les données de terrain montrent que, de façon générale, les cheptels sont soit indemnes, soit rapidement infectés à plus de 20-25 % en raison de la diffusibilité de l'agent pathogène.

Les paramètres suivants peuvent alors être retenus pour le sondage :

. Choix des cheptels par tirage au sort grâce à une table de nombres au hasard ;

. Pour des raisons de coût, on ne cherche pas une précision superflue ; on se contente d'une précision relative de 33 %.

Le nombre de cheptels à contrôler est alors trouvé dans le tableau III à l'intersection de la colonne 10 % (taux d'infection) et de la ligne 10.000 (cheptels) ; n = 320.

. Enfin, dans chaque cheptel, on prélèvera un nombre de sérums tel que défini au tableau V à l'intersection de la ligne correspondant au nombre d'animaux et de la colonne 20 % (taux d'infection des animaux choisi comme seuil).

CONCLUSION

Il est souhaitable de concevoir un sondage en définissant au préalable ses objectifs et en réunissant les éléments d'observation permettant d'avoir uneidée préalable de la situation sanitaire que l'on cherche à objectiver. Ces éléments permettent de définir les paramètres destinés à l'élaboration du protocole puis de calculer son coût financier. L'ensemble permet donc de reconsidérer éventuellement l'opportunité d'un tel sondage si les moyens financiers disponibles ne permettent pas, en particulier, d'atteindre la précision souhaitée.

BIBLIOGRAPHIE SOMMAIRE

DEROO M., DUSSAIX A.M.- Pratique et analyse des enquêtes par sondage (1980), Presses Universitaires de France (Paris 6ème).

ROUQUETTE C., SCHWARTZ D.- Méthodes en épidémiologie (1970), Editions Flammarion (Paris 6ème).

*
* *